

*In the Matter of Edward Brownlee,  
et al., Fire Fighter (M9999D) and (M9999E)*  
DOP Docket No. 2004-3328  
**(Merit System Board, decided September 22, 2004)**

Edward Brownlee, Antonio Capote, Anthony Cirelli, Christopher Curko, Joseph Dagato, Jorge Dominguez, Norman Figueroa III, Kevin Ingraham, Donald Leyhan, Brian Neill, Frank Pepe, Chris Reilly, Joseph Straile, and Ronald Winkelman represented by Craig Gumpel, Esq., appeal their scores for the written teamwork portion of the examination for Fire Fighter (M9999D) and/or for (M9999E). It is noted that appellants failed this portion of the examination. It is noted that these appeals were consolidated due to common issues presented by the appellants.

The examinations for M9999D and for M9999E were identical. However, the examination for M9999D was announced first, had a closing date of September 30, 2002, and consisted of announcements for non-consent decree jurisdictions only. The examination for M9999E had a closing date of April 30, 2003, and included the 12 consent decree cities along with other non-consent decree cities not listed for M9999D. Each examination consisted of a written cognitive portion, written teamwork portion and a physical performance portion. For M9999D, 4864 candidates were admitted to the examination, while for M9999E, 8549 candidates were admitted. The written teamwork portion of the test for M9999E was administered on several days in September 2003, while the written teamwork portion of the test for M9999D was administered on October 25, 2003. Approximately 120 individuals took make-up exams on later dates.

Also, some candidates that were admitted to M9999D, such as Messrs. Brownlee, Curko, Reilly and Zingaro were also admitted to the examination for M9999E. The candidates for the examination for M9999D were scheduled to take the written teamwork portion of the examination on October 25, 2003. Since candidates for both examinations had already taken this portion of the examination in September, these candidates were sent a form letter dated October 17, 2003. This letter stated, "Our records indicate you recently participated in the examination process for Fire Fighter, symbol M9999E during September, 2003. One of the examination parts for this test consisted of a Teamwork (Biodata) section. Our records further indicate you also applied for Fire Fighter, symbol M9999D, which also has the same Teamwork section as part of the examination process. As a result of your participation in the September 2003 testing process, your Teamwork score will be applied to Fire Fighter, symbol M9999D."

Candidate scores on the teamwork component ranged from 18 to 82 for M9999D, and from 12 to 87 for M9999E, with 46 being required to pass either test. Appellants all received scores of less than 46.

On appeal, appellants contend that the test is invalid and doesn't measure the merit and fitness of a candidate. They also state that it was not fair and competitive as it was the same examination as the one administered in 1999, which some candidates had

taken before, and therefore the Department did not employ precautionary measures to ensure the integrity of the examination.

Appellants then provide their backgrounds and teamwork activities, such as team sports, military service, experience in various professions, experience as a provisional Fire Fighter, training, fire fighting family members, knowledge of the jurisdiction for which the individual applied, serving as a volunteer firefighter, other voluntary service, certification as an Emergency Medical Technician, personal attributes such as honesty, and descriptions of incidents such as assisting in emergencies. Some appellants state that they passed the prior examination administration for M9999A.

The written teamwork portion was first administered in November 1999 to approximately 10,100 candidates as part of the examination for Fire Fighter (M9999A). The examination was administered under high scrutiny as consent decree cities were involved in that announcement. More than 500 appeals of test validity were received for that administration. In its decision, *In the Matter of Michael Abrams, et al., Fire Fighter (M9999A)*, (MSB, decided December 18, 2001), the Board discussed the validity of the examination at length and concluded that the teamwork component is a valid predictor of job success. In a related matter, 33 appellants were represented by Craig S. Gumpel, Esq., in the Board decision *In the Matter of Daniel Antonelli, et al., Fire Fighter (M9999A)*, (MSB, decided December 18, 2001), *aff'd*, *In the Matter of Steven T. Dill et al.*, Docket No. A-2671-01T2 (App. Div., September 2, 2004). In that decision, appellants raised complaints similar to the current appellants. They contended that the test was invalid or scored incorrectly, not job-related, discriminatory, poorly drawn, incomplete, inappropriate, and not clearly explained. The Board denied their appeals and the court affirmed the Board's decision.

In its decision *Antonelli v. New Jersey*, 310 F. Supp. 2d 700 (D.N.J., March 31, 2004), the United States District Court dismissed the challenges to the entry level Fire Fighter examination. Two groups of individual plaintiffs (24 in one group and 3 in the other group) and the FMBA challenged the exam, claiming that it violated their rights to due process and equal protection under the United States Constitution, their rights secured by 42 U.S.C. §1983, their rights under the New Jersey Constitution and civil service law, and the 1980 Consent Decree. In particular, they alleged that the biodata, or "teamwork" component of the exam was not job-related, that the DOP scored the exam in a race-conscious manner to increase the number of minority candidates on the list, that the DOP violated a previous court order by according the "teamwork" component more than a one-third weight, and that the DOP denied the plaintiffs sufficient access to test materials in order to challenge their failing scores. Defendants State of New Jersey and the United States brought a motion for summary judgment, which was granted by the court in an opinion issued by Judge William Walls on March 31, 2004.

In his opinion, Judge Walls dismissed the FMBA's complaint for lack of standing. He also held that the individual plaintiffs did not have standing to enforce the Consent Decree, because they were not parties to that matter. Further, he held that the plaintiff's remaining claims, with one exception, were barred by New Jersey's sovereign immunity

pursuant to the Eleventh Amendment. With regard to the one exception, the Section 1983 claim, Judge Wells held that the plaintiffs failed to show that the DOP officials acted under color of state law to deprive them of a federal constitutional right. Specifically, he found that there was no denial of equal protection since the exam was not discriminatory in its content, scoring or administration. Moreover, there was no denial of due process in their limited access to testing materials, since the plaintiffs did not have a property right in employment as firefighters. Rather, they only had a desire to become firefighters, and even if they had passed the exam and become candidates, they did not have a vested right to appointment.

The test in question here was administered pursuant to the requirements of a federal Consent Decree which governs firefighter testing in New Jersey. A history of this decree was explained in *IMO Michael Abrams, et al.* and need not be repeated herein. It is noted that the decree permitted the Department of Justice (DOJ) to object at any point in the selection process, and the federal court retains jurisdiction over all matters covered by the decree and hears all disputes. Pursuant to the requirements of a supplemental consent decree, the Department developed the Fire Fighter (M9999A) examination. As explained in *Abrams, et al.*, to ensure that a high quality examination was developed, the department committed more than 21,000 hours of staff time to the design and development of the firefighter examination, at a cost exceeding \$3 million. The development process began in 1990 when the Department initiated a job analysis of the entry-level firefighter position, which was completed in October 1995 after five years of extensive research. The Department set forth its research, data and conclusions in the “Fire Fighter Job Analysis Study” (“Job Analysis”). The DOJ posed no objection to the Job Analysis.

Based upon the results of the Job Analysis, three distinct areas of abilities were identified as equally important to performing the job of a firefighter: cognitive, physical performance and a third area, which was neither cognitive nor physical in nature, referred to as “teamwork” abilities by the Department. Teamwork abilities include working with others as a team; getting along well with people of diverse backgrounds and personalities; accepting and following orders; showing respect for the work contributions of others; having integrity; being dependable; and having the willingness to share information with and help others.

The Department then contracted to develop an examination which would cover the three test component areas, cognitive, physical performance, and teamwork. Since cognitive and physical testing instruments are not designed to measure an individual’s teamwork ability, a biodata test was developed. This selection methodology, which is also referred to as a biographical inventory, is commonly used in entry-level hiring as an alternative or supplement to cognitive testing because it is able to predict aspects of job performance that are not predicted by cognitive measures.<sup>1</sup> Biodata focuses on a person’s past and current history by identifying those life experiences and behaviors that are similar to those of successful incumbents in the job in question.

---

<sup>1</sup> Crafts J., Using biodata as a selection instrument. *ERIC Clearinghouse on Tests Measurement and Evaluation*. ERIC Identifier ED338702 (1991), p.1

Within the field of personnel selection, it is widely accepted that biodata instruments are valid and reliable predictors of job performance.<sup>2</sup> The scholarly literature indicates that these instruments have been shown to be valid predictors of job performance in numerous settings and for a wide range of criteria such as job performance, training success, team performance, job adjustment, job turnover, job choice, etc.<sup>3</sup> Compared to other selection measures, biodata validity coefficients have been found to be impressive, often with demonstrably higher validities than other selection procedures.<sup>4</sup>

A number of major employers, including American Express, Circuit City, Eli Lilly, GTE, IBM and Prudential, use biodata tests to select job candidates. A number of public employers also use such tests. Several use biodata instruments for firefighter candidates, such as Los Angeles, Tucson and the State of Kansas, and several use it for law enforcement titles, such as the FBI, the U.S. Army and the states of Kansas, Iowa and Louisiana.

A contractor, Dr. Terry Mitchell, of ESelex.com, San Diego, California, developed this type of test for the firefighter title by collecting extensive data on firefighter performance, beginning with a process known as the multiple criterion antecedent procedure (MCAP). The first step in MCAP involved the review of the firefighter job analysis to identify critical firefighter tasks and KSAs. The next step was to analyze the underlying worker characteristics associated with these tasks and KSAs. This was accomplished by interviewing members of the New Jersey fire service about these worker characteristics.

The final step of the MCAP identified the aspects of life history underlying the development of worker characteristics. For example, incumbent firefighters were asked to identify and discuss matters such as events in their childhood, family, school, early adulthood, or work experience, which may have contributed to the development of relevant worker characteristics. The resulting life history variables, developmental experiences and behavioral incidents became the foundation for constructing biodata items. In this way, the biodata items were directly and rationally linked to the critical tasks and KSAs of the firefighter job.

These biodata items were used to create a biodata survey, which was then administered to a sample of approximately 1000 firefighters. The supervisors of these firefighters – both immediate supervisors and those at the next higher supervisory level – were identified and asked to complete a standardized evaluation form to rate the performance of the firefighters in the sample with respect to tasks, KSAs, and characteristics identified in the firefighter job analysis and the MCAP.

---

<sup>2</sup> Baird, 1976; Reilly & Chao, 1982) [In Smith, H.A., & Pratt, D., The use of biodata in admissions to teacher education. *Journal of Teacher Education*, 47 (1996), p.44]

<sup>3</sup> [Hunter & Hunter, 1984; Schmitt, Gooding, Noe, & Kirsh, 1984] [In Melamed T., Use of biodata for predicting academic success over thirty years. *Psychological Reports*, 71 (1992), p. 31]

<sup>4</sup> Elkins, T. J. & Phillips, J.S., Job context, selection decision outcome, and the perceived fairness of selections test: Biodata as an illustrative case. *Journal of Applied Psychology*, 85 (2000), p. 103

For example, supervisors were asked to rate incumbent firefighters on how they got along with individuals with different backgrounds or with different personalities. This produced 730 matched sets of data. That is, for every one of the 730 firefighters who completed the biodata survey, there was performance evaluation data available on that firefighter based on the supervisor's input. The comparison of these firefighter evaluations with the collected biodata from the respective firefighters provided the basis for determining those items that predict job performance.

According to the Uniform Guidelines and professional testing standards, a test must be validated to ensure the test items relate to potential job performance. An employment test is valid if it accurately predicts success in the critical job aspects being measured. American Education Research Association, American Psychiatric Association and National Council on Measurement in Education (1999), *Standards for Education and Psychological Testing*. The Uniform Guidelines strongly urge the use of validation methodology, and the DOJ insisted here that the biodata instrument be validated by this method. A criterion validity study attempts to demonstrate a statistically significant relationship between a predictor (in this case, biodata items), and a criterion measure (here, supervisory ratings).

As part of this validity procedure, the 730 matched sets of data were divided into two equal parts. The first group of 365 (randomly selected), the derivation sample, was used by Dr. Mitchell to develop biodata predictor scales, i.e. the multiple choice response options and the weights of these options. These included the teamwork predictor scale. The teamwork predictor scale constitutes those biodata items (questions, options and option weighting) that are related to teamwork and that best differentiate among incumbent firefighters in terms of their teamwork experience. This was possible because the data showed a relationship between incumbent firefighter responses to the teamwork items and the supervisory ratings of those firefighters with regard to their job performance on teamwork-related behaviors. That is, firefighters who received high teamwork ratings from their supervisors answered questions differently than firefighters with lower teamwork ratings, indicating that certain teamwork-related responses on the biodata survey successfully predicted their present teamwork performance. Thus, the scale identified scoring criteria to predict the teamwork-related behaviors of firefighters.

These profiles were applied to the biodata responses obtained from the second group of 365 matched sets, the cross-validation sample, to confirm that the criteria were consistent and reliable. This cross-validation process revealed a statistically-significant correlation between the teamwork biodata questions and the teamwork performance criterion factor, the measure of incumbent firefighter teamwork abilities. Thus, there is clear evidence of the criterion-related validity of the biodata predictors. That is, the correlation between the teamwork performance criterion factor and the teamwork biodata questions demonstrates that the teamwork component questions predict job success in the area of teamwork.<sup>5</sup>

---

<sup>5</sup> It is noted that the data also show evidence of construct validity for the teamwork component. See Federal Uniform Guidelines on Employee Selection, Sec. 14D(2).

Dr. Mitchell's data and analysis further supported the conclusion that the critical aspects of the firefighter job are divided into cognitive, teamwork and physical performance. This analysis also determined that these three areas of performance are of equal importance in contributing to overall job performance. Accordingly, the Department determined that the abilities to be measured in the teamwork component constitute one-third of the necessary overall abilities.

Each aspect in the development of the teamwork test component was exhaustively reviewed by the DOJ and its experts. The DOJ determined that Dr. Mitchell's test was properly designed to measure the critical teamwork aspects of the job of firefighter. In fact, DOJ took the position that the teamwork biodata test should be the only component used for ranking candidates, with the cognitive and physical test portions being pass-fail. The Department of Personnel disagreed that the teamwork test should be the sole determinant of rank; it took the position that the teamwork, physical and cognitive aspects of the job were of equal importance, and therefore each of these components should count as one-third of a candidate's total score.

The Department and DOJ were unable to reach an agreement over the weighting and scoring of the test, due to DOJ's insistence that the biodata questions had to be the sole ranking device. In view of this impasse, the Department filed a motion in federal court for an order establishing each of the three test components as one-third of the candidate's total score. This motion was supported by the two major unions which represent New Jersey firefighters, the FMBA and the IAFF, and opposed by DOJ. On June 15, 1999, United States District Judge Nicholas H. Politan granted the Department's motion. The judge determined that the test "measures three significant functions that a person should have: cognitive, physical, and bio teamwork.... What you're really measuring is the person, himself, and whether he's a good team player." (Tr. at 62). The judge authorized the Department to announce the firefighter test with each section afforded a scoring weight of one-third. The judge further ordered the State and DOJ to discuss the use of the biodata test in view of his determination.

The Department announced the Fire Fighter (M9999A) examination on June 21, 1999. Shortly after June 1999, in accordance with the court's order, the Department reviewed the use of the biodata portion of the test. This review showed that the entire biodata test should not be scored, because, according to the test developer, a number of questions on it measured the physical or cognitive skills of the candidate, rather than the teamwork skills. For example, Dr. Mitchell's analysis showed that some of the questions regarding educational background were better predictors of a candidate's cognitive ability than of teamwork ability. Scoring these questions would have made the biodata portion of the test actually account for more than one-third of the total score, in violation of the court's order. Further, the Department determined that it was more appropriate to use a performance test than a written test to measure physical abilities. For example, the Department felt that having a candidate actually climb a ladder or drag a 160 pound dummy was a better way to measure physical performance than using a "biodata" questionnaire. The Department felt that it was more of a direct measure of these physical

abilities and, therefore, it would be more acceptable to the candidate population. Acceptance of the testing method by candidates was a critical factor in the administration of the new examination. Thus, to avoid using the biodata test to partially determine physical and cognitive performance, the Department decided, in September 1999, that only the teamwork questions would be scored. However, it decided to administer the entire 141-question test to collect data for research purposes. Further, it is more difficult for candidates to discern which of the biodata items were teamwork questions as this might, in turn, allow them to discern a pattern which could allow the identification of the desired answers to the those questions. This is intended to address critical test security concerns.

For M9999A, the Department scored the cognitive and physical tests and Dr. Mitchell graded the teamwork test. This procedure is the same for M9999D and M9999E. The raw scores for each examination component were standardized (i.e., converted to Z scores). The Department set passing points for each component of the examination, based on analysis of the mean scores, consistent with professional testing standards.

For M9999A, the passing scores established by the Department resulted in no adverse impact. These passing points were reviewed and deemed appropriate by an expert retained by the Department, Dr. Paul Sackett. The DOJ then reviewed the passing points along with the overall test results, and had no objection. The passing point remains the same, 46, for M9999D and M9999E.

## **CONCLUSION**

These appeals have been consolidated because they all involve the same claims. With regard to the teamwork test, the appellants all contend that the test was invalid because they failed, even though they are team players. In support of this argument, they rely on various aspects of their backgrounds which demonstrate, appellants say, that they should have received higher teamwork scores. Appellants cite the following examples to show their teamwork ability: participation in various sports; knowledge of the jurisdictions in which they seek employment; employment in various professions; volunteer firefighting experience; provisional experience; family members with fire fighting experience; Emergency Medical Technician certification; and attendance at classes or training.

Initially, it is noted that some appellants submitted certificates, licenses, awards, letters of recommendation, and explanations of their life experiences. These documents cannot be considered as they were not part of the examination. All candidates were rated in the same way, according to equal standards; that is, all scores were calculated based exclusively on the answers given by candidates in response to the questions on the test. Accordingly, the materials submitted cannot affect the outcome of the exam. All candidates took the same test and were rated according to the multiple choice answers they selected, which had previously been validated as job related by Dr. Mitchell. It is noted that the written teamwork portion of this examination was a multiple-choice and "fill in the blank" format, but only the multiple-choice questions were scored. A small

portion of questions asked for information in short-answer or “fill in the blank” format, but these were not used in scoring. The answers to those questions are used only for verification purposes.

With regard to appellants’ reliance on the above-listed life experiences, the fact that appellants had these experiences does not demonstrate the invalidity of the teamwork examination. Appellants all challenge the test on the basis that it was not valid, i.e., that it did not select the individuals with the best teamwork ability. The claim of invalidity rests exclusively on the fact that appellants failed despite having taken fire-related courses and having participated in activities like fire fighting and team sports. Based on this premise, appellants conclude that “the most qualified candidates for the position were excluded through the use of biodata.” In other words, appellants assume that the candidates who passed the test did not have these types of experiences. However, this is not the case.

The record shows that successful candidates had fire fighting, team sports and military experience, as well as fire-related education. An analysis of all candidate responses for M9999A reveals that the individuals who passed the biodata component had the same or greater levels of experience as the failing candidates on the same type of backgrounds relied upon by appellants. For example, 34% of the candidates who passed the biodata test, and 34% of those who failed, stated that they had participated in at least 3 or 4 years on a varsity team sport in high school, and 52% of both groups indicated participation in band, plays, clubs and sports in high school. With regard to fire science education, more of the passers took at least one such class (17% v. 12%), and slightly more of the passers took 5 or more such courses (4.6% v. 3.3%). In addition, substantially more of the passers had fire fighting experience (25% v. 16.5%) and military experience (16.4% v. 9.9%).

Thus, the backgrounds and experiences identified by the appellants for M9999A were not exclusive to either the passing or failing candidates. The applicants who passed the Teamwork component did display different response patterns than applicants who failed. That is, based on their responses to the Teamwork-related questions, passing candidates are predicted to demonstrate higher levels of Teamwork-related abilities (such as “showing respect for the work contributions of others,” “willingness to share information with and help others,” etc.) than applicants who failed the Teamwork component. This is based on the validation study and the expert opinion of Dr. Mitchell. This analysis was performed for M9999A. Appellants’ claim that the biodata test did not select the best candidates is not supported by any evidence. Even if the appellants’ selection of experiences were considered to be valid criteria to predict job success, it could still be said that candidates on the eligibility lists are well-qualified and highly credentialed under that criteria. In fact, in terms of some of the appellants’ criteria, a greater percentage of passing candidates as a group indicated that they possessed those very experiences (e.g., fire fighting and military experience) than the failing candidates. Since it was found that the successful candidates had fire fighting, team sports and military experience, as well as fire-related education, these results do not warrant a similar analysis for M9999D and M9999E.



Appellants have not presented any evidence to support their contention that the test is not valid as that term is defined by the governing consent decree and legal requirements. The decree required that the teamwork examination be validated according to professional testing standards and federal law. The Uniform Guidelines expressly describe how tests must be validated. The Department's expert, Dr. Mitchell, followed these standards and demonstrated through an appropriate criterion validity study that the teamwork test was valid. The Justice Department and its experts exhaustively reviewed the test validation and determined that it was performed in accordance with these standards; and the federal court approved the overall examination and the eligibility lists based on it. The teamwork portion of the examinations for M9999D and M9999E contained the same questions, although the order in which they were given was scrambled.

Regarding the fairness of the administration of the examination, appellants rely on *In the Matter of Police Sergeant* (PM3776V), *City of Paterson*, 176 N.J. 49, 52 (2003), wherein the Court decided that the Department should administer make-up civil service examinations which contain substantially different or entirely different questions from those used in the original examination. In that case, make-up examinations were given for a promotional Police Sergeant examination after the Department was made aware of breaches in security, specifically, when 65 of the 71 multiple choice questions had been typed out and disseminated throughout the Paterson Police Department. The court found that the Department's practice of administering identical exams to original and make-up candidates in the same testing cycle was not *per se* unconstitutional, but that petitioners' right to a fair and competitive examination was violated in light of direct evidence that information about the content of the examination was available to applicants who had not yet taken the exam. See *N.J.S.A.* Const. Art. 7, § 1, par. 2.

Appellants argue that giving the same examination where there was evidence that information about the content of the examination was available to applicants who had not yet taken the exam violated *N.J.S.A.* Const. Art. 7, § 1, par. 2. First, the current examination is not a make-up examination and is not in the same testing cycle. Secondly, in the instant matter, no evidence of information about the content of the examination was available to applicants who had not yet taken the exam has been discovered by the Department, and appellants have not provided any proof that any candidate provided information about the content of the examination to applicants who had not yet taken the exam. In fact, counsel for the current appellants pleaded for access to test materials, including the questions and answer key, in order to challenge failing scores. See *Antonelli*, 310 F. Supp. 2d 700, 709.

Appellants also rely on *Rox v. Department of Civil Service* 141 N.J. Super. 463 (App. Div. 1976), wherein the Appellate Division invalidated a promotional examination after it determined that the subjective nature of the exam rendered it "noncompetitive and illegal." *Id.* at 467. The court stated that it was important that the Merit System Board adopt "adequate and meaningful precautionary measures" to assure examination security. *Id.* at 469-70. They state that the facts in *Rox* are nearly identical to the subject

examination in that numerous applicants had foreknowledge of the exam as they participated in the 1999 Fire Fighter as well.

In reply, the facts in *Rox* are clearly dissimilar. *Rox* involved a Police Captain promotional examination, while the subject examination is open-competitive. The Police Captain examination was an oral examination requiring specific answers to be given, while the component under appeal is a written teamwork examination which requires a candidate to answer questions about his or her own background, history and judgments. No candidate has been provided with the scoring key, or a review of the scoring key, for the written teamwork component of the examination. There are no specifically correct answers to this non-cognitive component. As stated on the sheet, "Scoring of the teamwork component of the fire fighter exam" given to candidates, options to each question were weighted in scoring. That is, options indicating strong teamwork skills were given more credit than options indicating weak or no teamwork skills. Even if the questions and the multiple options were memorized and copied down, candidates were not provided the weighting of the options. In addition, questions related to cognitive and physical skills were not graded, and candidates were not told which questions were scored.

Appellants have presented no evidence that contradicts the validity and job-relatedness of the biodata test. They simply state that a job applicant can be successful without having the same work and life experiences as present firefighters. This unsupported assumption does not overcome the clear evidence of test validity described above.

Appellants argue that there was a disparity in treatment of candidates who had applied for consent decree jurisdictions and those who had applied for non-consent decree jurisdictions, stating that those who applied for consent decree jurisdictions were not permitted to take the physical component of the exam if they failed the teamwork portion. In reply, as stated above, M9999D was announced first with a closing date of September 30, 2002, and consisted of announcements for non-consent decree jurisdictions only. The examination for M9999E had a closing date of April 30, 2003, and included the 12 consent decree cities along with other non-consent decree cities not listed for M9999D. The contract for the written teamwork portion was not finalized until the summer of 2003. M9999D was announced eight months prior to M9999E, and there was no reason to delay the examination of the written cognitive portion and the physical portion until the contract on the written teamwork portion was finalized. As such, the written cognitive portion was administered and those candidates who passed were given the physical performance portion. Candidates who passed that were scheduled for the written teamwork portion. For M9999E, the contract for the written teamwork component had been finalized by September 2003. As such, all candidates were given both written components at the same time. Candidates who passed BOTH written portions were scheduled for the physical performance portion. All candidates had to pass all three components of the examination in order to pass the examination, and appellants provide no arguments as to why the order of the administration of the components would constitute disparate treatment. Also, candidates were tested based on which examination

they applied for, and M9999E tested both consent decree and non-consent decree jurisdictions.

Appellants request a review of the teamwork test questions and answer key, as well as all other test material. Such review cannot be permitted, in order to maintain test security. The Commissioner of Personnel is authorized to preclude candidate review of exam materials. *N.J.A.C. 4A:4-6.4(e)*. In this instance, candidates were given an explanation of the teamwork examination and its scoring, but were not permitted to review the examination or the answer key. Such action was appropriate to ensure that the test questions could be used again.

Precluding test review is crucial under the special circumstances of this matter. The Department has invested significant amounts of time and money in producing the firefighter test. If the test questions and answers become known to candidates, it will render the test useless. As discussed, it took nearly 20 years, several million dollars, and the time of over 700 firefighters and their supervisors to produce a firefighter test that satisfied the requirements of the consent decree and the Uniform Guidelines. In view of the large number of appellants, there is a real risk that some of these individuals would share information about the test with other candidates, or would use such information themselves the next time the test is given.

The need to preserve test security is replete throughout the psychometric literature. For example, according to the *American Psychological Association*, "Public exposure of test materials should be evaluated in light of the ethical obligations of professionals to protect the value of secure tests whose psychometric integrity depends upon the test taker not having prior access to test materials."<sup>6</sup> Similarly, "Disclosure of secure testing materials to professionally unqualified persons may decrease the test's validity. Availability of test items to an unqualified person can not only render the test invalid for any future use for that individual, but also can jeopardize the security and integrity of the test for other persons who may be exposed to test items and responses. Such release can impose very concrete harm to the general public in the loss of effective assessment tools."<sup>7</sup>

Appellants have challenged the overall validity of the test. Even if the appellants had access to the test questions, they would not be able to prove that the test is invalid. The test itself has already been validated through appropriate psychometric concepts, consistent with the Uniform Guidelines, and this validation has been reviewed by both Department consultants and the DOJ consultants, and approved by the Court. Permitting appellants access to review the questions and answers then, would serve no purpose, except to expose the test and render its future use moot. Also, full access to the scoring would impair the Department's ability to contract with private testing firms to provide this service.

---

<sup>6</sup> American Psychological Association, Inc. (1999), p. 1078

<sup>7</sup> Blackwell, Autry & Guglielmo (2001), p. 164

*Brady v. Department of Personnel*, 149 N.J. 244 (1997), does not require a different result. In *Brady*, a police sergeant appealed his grade on the written essay portion of a police promotional exam. He contended that he should have been given higher scores for his answers. Although Brady was permitted to review some test materials, including brief summaries of the test questions and brief comments by the grader of his test, he complained that he needed access to the actual test questions and answer key in order to challenge his scores. The Supreme Court upheld the Department of Personnel's restrictions on reviewing these materials, noting that "full disclosure would wreak havoc with the Department's legitimate efforts to maintain security." *Id.* at 261. The Court rejected Brady's argument that he was entitled to "full access to the testing materials as a basis for challenging and obtaining broad judicial review of the accuracy of the agency's scoring of the examinations." *Id.* at 260. The Court based this conclusion on the principle that courts will not review civil service tests to determine whether questions "were well or poorly answered," but may only determine "whether the testing and grading were clearly arbitrary." *Id.* at 258.

Although the Court in *Brady* suggested that the denial of access to all test materials would generally be improper, *Id.* at 262, this statement does not apply to the present case, in which there are exceptional circumstances. *Brady* applies only to the situation where the candidate claims that the Department has misgraded his answers to certain questions. In such a case, *Brady* holds that the candidate is entitled to evidence that enables him to assess whether the Department acted arbitrarily in scoring his test. In contrast, the appellants do not claim that their answers to *certain* questions were incorrectly graded. Instead, they attack the overall validity of the examination, contending that the most qualified candidates for the position were excluded through the use of biodata, or that they were incorrectly scored somehow. This claim does not depend upon access to the test questions and answer key. As discussed above, the undisputed test results and validation study contradict appellants' allegation of test invalidity. Also, the teamwork questions are non-cognitive and require a candidate to answer questions about his or her own background, history and judgments. There are no specifically "correct" answers: the options are weighted, and revealing this information would compromise the security of the test and give an unfair advantage upon re-examination. See *In the Matter of Steven Dill, et al.*, Docket Nos. A-2674-01T2, A-2675-01T2, A-2678-01T2 (App. Div. September 2, 2004).

Moreover, *Brady* did not involve an examination developed pursuant to a federal consent decree. The Court's concern in *Brady* was to ensure that the Department did not conduct testing without any accountability. Such accountability is present here. The validity of the biodata test was reviewed by the Justice Department, and was subject to the oversight of the federal court. In view of the history of the extraordinary review given this test under the consent decree, there is no basis to infer that the test and its grading were arbitrary.

N.J.A.C. 4A:4-6.4(e) states that, "in order to maintain the security of the examination process, the Commissioner may, on a particular examination, modify or eliminate the review of the examination questions and answers." As noted in *Brady*,

legislation specifically *requires* the Department to “provide for the security of the examination process,” *Id.* at 259. See *N.J.S.A.* 11A:4-1(c). The Department must reasonably balance test security with candidate interests. In this instance, examination reviews were eliminated as the security of the examination outweighs the necessity of candidates to observe examination scoring in order to verify their scores. Appellants have not provided specific allegation of arbitrariness or discrimination, or a *prima facie* showing of more than the mere possibility of misgrading.

For M9999A, as an additional check on the grading, the Department asked the test developer to rescore the teamwork component answer sheets. This review confirmed the accuracy of the scores. As a check was performed for thousands of candidates for M9999A and no errors were found, a rescoring of M9999D and M9999E, which would cost in excess of \$25,000, is not warranted.

Appellants also broadly state that unidentified teamwork test questions were not job-related or ambiguous. Appellants have provided no support for this contention. In view of the job analysis and validation study, together with the exhaustive review accorded this test under the consent decrees, there is no basis to conclude that the test questions were not job-related or were otherwise improper.

Regarding admission to both examinations, it was decided that the same examination could be administered as the teamwork portion was based on life experience. A person’s life experiences do not change appreciably over the course of two months that it would affect the way the candidate would respond to the questions. The questions were the same for both tests, except they were scrambled. A candidate would not answer the questions differently on different test dates. Historically, from a research perspective, it would take a year or more before a candidate’s score would change. Because of that, the vender agreed to score one test per person per year. At this juncture, enough time has passed since the examination was administered that if appellants were allowed to retake the examination now, it would be an unfair testing situation, as they would have the advantage of more life experience than the other candidates who took the examination in either September or October 2003.

Lastly, Mr. O’Neill states that he has a medical condition which would require ADA assistance. He states that the application instructions did not clearly state that he had to request assistance by checking box 13, and therefore, he did not get the assistance he required.

Regarding ADA assistance, Mr. Neill was not treated unfairly in that he was tested under the same conditions as all other candidates who did not request ADA assistance. If appellant thought he would need assistance, he would have had to ask for it at some point in time prior to the administration of the examination. Given appellant’s medical condition as he has described it, and the information and sample questions provided to him in the study guide, it is unclear why he did not request ADA assistance prior to the administration of the examination, even if he had not seen the box on the application. Instead, appellant filed an appeal on March 30, 2004, over four months after

the administration of the examination and after receiving the examination results, claiming that he did not know that that he could ask for assistance.

The directions on the application told candidates to check box #13 on the front of the application “if you would like to be contacted regarding an auxiliary aid or a reasonable accommodation in taking this examination in accordance with the Americans with Disabilities Act.” Under “Job Application Process” on the Department of Personnel website, under “Additional Information” the following information is listed regarding ADA assistance:

**Americans with Disabilities Act ( ADA ) Assistance**

Persons with disabilities are eligible to register for and take examinations as long as they satisfy the job requirements for educational background, employment experience, skills, licenses, and any other qualification standards that are job related, and are able to perform those tasks that are essential to the job, with or without reasonable accommodation. If the accommodation cannot be made because it would cause the employer undue hardship, such persons may not be eligible.

If you have a serious visual impairment (including blindness) or any other significant disability which makes you unable to take the required examination in the traditional manner, the Department of Personnel will make arrangements to reasonably accommodate your need. Our test facilities are wheelchair accessible.

You will be asked to advise us if you have a need for reasonable accommodation as noted above. It is recommended that you advise us of your need as quickly as possible to allow sufficient time for us to arrange the needed accommodation.

You will be asked to provide official documentation of your disability.

To request an accommodation, call the Department of Personnel at (609) 292-4144.

All candidates had to take all parts of the examination. If appellant thought he needed assistance on the examination, he was told he had to indicate the need for assistance on his application, and he could have called to verify this information. As appellant requested this information after he took the examination and received his score, his request for reasonable accommodation is untimely. See *In the Matter of Leonard S. Cohen*, Docket No. A-5189-99T5 (App. Div. September 25, 2001).

As to the merits of the appeal, when a candidate requests reasonable accommodation or ADA assistance by checking box 13 of the application, the Division of Selection Services sends the candidate a Special Accommodation Request form which includes a list of general accommodations used on the first side of the form, and a section to be completed by a doctor or child study team on the second side of the form. Accommodations generally provided include readers, markers, interpreters, extra time, separate rooms, special parking, mobility assistance, wheelchair access, special seating, or a personal attendant provided by the candidate. Special accommodation would not include rewording or clarifying questions, as this would be considered assistance in determining the correct answer to questions. For Mr. Neill, a physician or child study

team has not provided support to appellant's claim of need for the ADA accommodation and a description of the special accommodation needed for this examination.

A thorough review of the record indicates that the determinations of the Division of Selection Services were proper and consistent with Department of Personnel regulations, and that appellants have not met their burden of proof in these matters.

## **ORDER**

Therefore, it is ordered that these appeals be denied.

This is the final administrative determination in this matter. Any further review should be pursued in a judicial forum.